XIV. SPEECH COMMUNICATION*

Prof. K. N. Stevens	Prof. J. M. Heinz	A. P. Paul
Prof. M. Halle	Dr. A. S. House	S. S. Reisman
Prof. J. B. Dennis	Jane B. Arnold	E. C. Whitman
	W. L. Henke	

A. STUDIES OF THE DYNAMICS OF SPEECH PRODUCTION

During the past year several cineradiographic films showing lateral views of the speech mechanism, together with sound recordings, were made in collaboration with the Speech Transmission Laboratory, Royal Institute of Technology, and the Wenner-Gren Research Laboratory, Norrtull's Hospital, in Stockholm, Sweden. The general objectives of this work have been outlined elsewhere.[1] Analysis of the films is now in progress, and some of this analysis is being carried out by the Speech Communication Group of the Research Laboratory of Electronics. This report presents examples of some preliminary results that are emerging from the analysis of the films.

The data presented here are taken from tracings made from each frame of the film throughout each of several utterances. The tracings represent midsagittal views of the outline of the vocal tract and of the surrounding structures. The time interval between frames on the film is 23 msec. Several measurements showing the motions of various components of the articulatory mechanism were made from the tracings. Attempts were made to improve the accuracy of the measurements by averaging groups of tracings representing contours that remain invariant in shape, such as the mandible, hard palate, and vertebrae, and by using plaster casts to define the contours of the teeth and hard palate.

The illustrative results represent data that provide measures of the activity of the tongue tip and the pharynx. Specifically, these measurements are: (a) the distance in the midsagittal plane from a point on the alveolar ridge to the tongue surface, measured along a fixed line drawn parallel to the posterior mandibular surface when the molars are approximated; and (b) the average of two measurements of the distance between the anterior surface of the vertebrae and the posterior tongue surface at the levels of the lower surfaces of the second and third cervical vertebrae, respectively. Results for four utterances by one talker are shown in Fig. XIV-1. The utterances are /hə'tɛ/, /hə'tɛt/, /hə'tɑ/, and /hə'tɑt/; they were selected because they illustrate the production of several allophones of one consonant phoneme, /t/, in several vowel environments and of two vowels, /ɛ/ and /ɑ/, in different consonantal environments.

Several features of the results shown in Fig. XIV-1, together with related data on

*This research was supported in part by the U. S. Air Force (Electronic Systems Division) under Contract AF 19(604)-6102; in part by the National Science Foundation (Grant G-16526); and in part by the National Institutes of Health (Grant MH-04737-03 and Grant NB-04332-01); additional support was received under NASA Grant NsG-496.

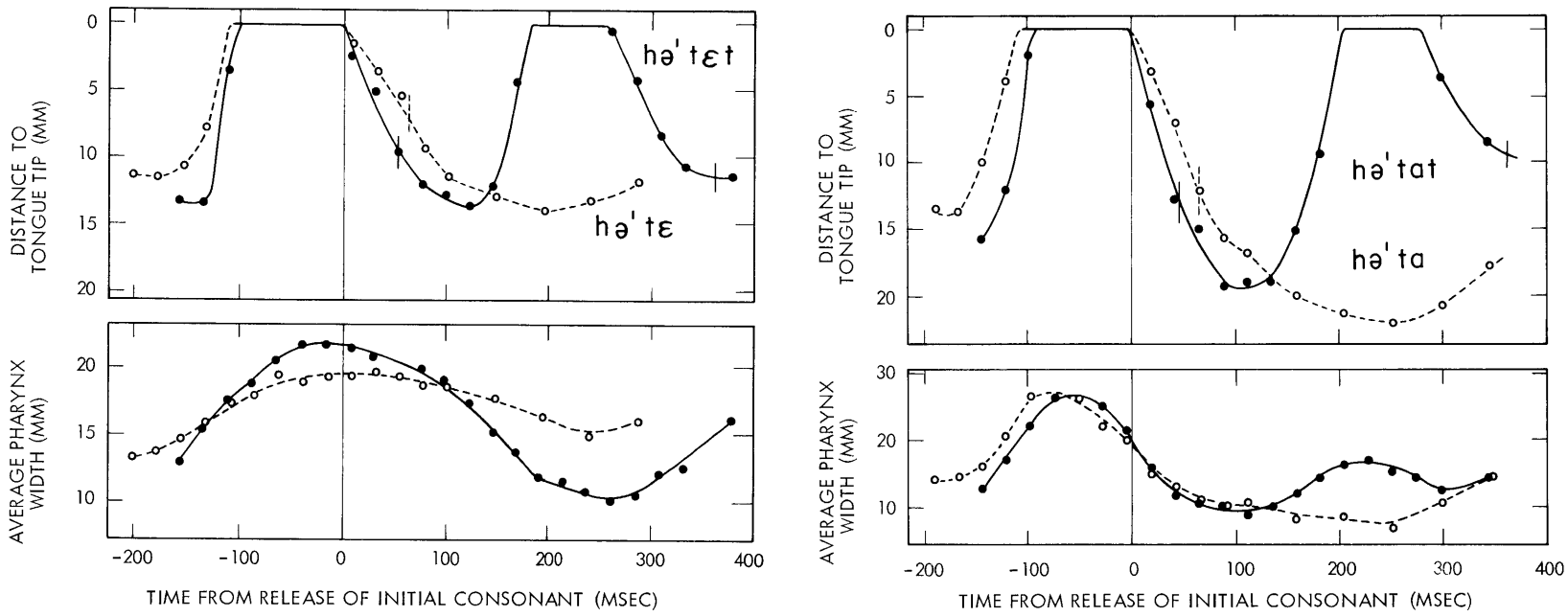


Fig. XIV-1. Graphs showing measures of displacement of tongue tip relative to alveolar ridge (upper curves) and pharynx width (lower curves) for four utterances as shown. The short vertical lines 40-60 msec to the right of the time of release of the consonants in the upper graphs represent the times at which aspiration terminates, as determined from spectrograms of the utterances.

other measurements for the same utterances, can be used to demonstrate some of the contextual effects that influence the articulatory activity associated with a given phoneme. The findings must, of course, be interpreted with caution, since they represent data for only a few utterances by one talker.

1. The rate of tongue-tip motion when the /t/ is exploded in intervocalic position is not the same for all versions of /t/, but depends on the following vowel and on the presence or absence of a final consonant in the stressed syllable. The rate of motion of the tongue tip following the /t/ explosion is greatest in /hə'tɑt/ and least in /hə'tɛ/. These results suggest that the rate of tongue-tip motion depends on how far the tongue tip has to go (greater for /tɑ/ than for /tɛ/), and on whether the tongue tip must return to the consonant position after it reaches the vowel "target" position.

2. The rate of motion of the tongue tip is considerably faster during the closing phase of /t/ than during the release. As a consequence, there is a marked asymmetry in the motion of the tongue tip during the syllables /tɑt/ and /tɛt/, where the closure following the vowel element occurs at a faster rate than the release initiating the vowel.

3. The aspiration phase during the release of the initial stop consonants in these utterances terminates only after the tongue tip has formed a relatively large opening. During this aspiration phase the vocal tract is presumably excited by noise caused by turbulence in the vicinity of the glottis rather than at the constriction formed by the tongue. In this interval the glottis is apparently becoming more constricted in preparation for the initiation of vocal-fold vibrations.

4. The pharynx width during the generation of these stop consonants depends strongly on the vowel environment. In the syllable /hə'tɛt/, for example, the pharynx width during the closure for the initial /t/ preceding /ɛ/ is considerably greater than the width for the final /t/, which precedes silence.

5. It is possible to select arbitrarily a point within each vowel at which the vowel "target" configuration is assumed to be reached most closely, which corresponds to the time when the displacement of the tongue tip is maximal. At this instant of time there appears to be a slight undershoot in both tongue-tip displacement and pharynx width when the vowel is in the interconsonantal environment, compared with the situation in which there is no consonant termination on the syllable.

K. N. Stevens

References

1. K. N. Stevens and S. Öhman, Cineradiographic studies of speech, Speech Transmission Laboratory Quarterly Progress and Status Report, Royal Institute of Technology, Stockholm, Sweden, July 15, 1963, pp. 9-11.

B. DESIGN CONSIDERATIONS FOR AN IMPROVED VOCAL TRACT ANALOG

Work on the development of an improved dynamic transmission-line analog of the vocal tract has been previously reported by this group.[1,2] A section of acoustic transmission line approximated by electrical inductance and capacitance as shown in Fig. XIV-2a is modeled by the configuration of components in Fig. XIV-2b. Since it must be possible to vary the effective area of a section over a range of at least 100 to 1, control of the multipliers by the logarithm of the coefficients is convenient and desirable.

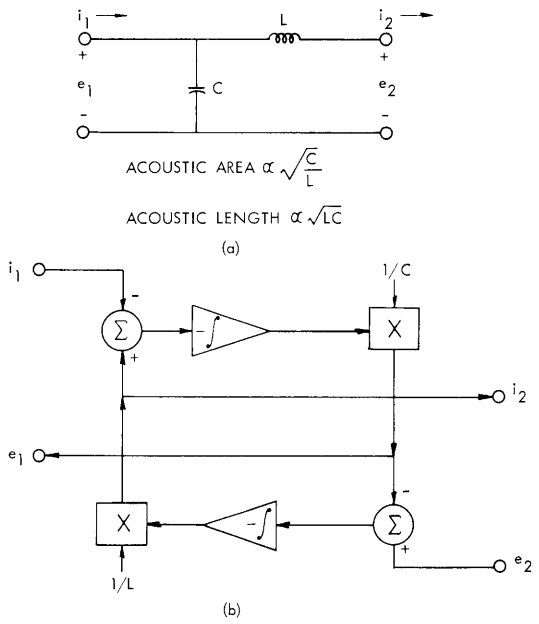


Fig. XIV-2. Electrical model of a section of acoustic transmission line and its representation by analog components.

Also, digital control of the multiplier coefficient is appropriate, since it is planned to operate the analog under the control of a digital computer. Therefore the implementation of the multiplying component in the form of a digitally controlled logarithmic attenuator is being studied. The principle of the attenuator is described by Fig. XIV-3. The attenuator is constructed of a cascade of switched sections having individual attenuations of 1/4 db, 1/2 db, 1 db, ..., 16 db, and 32 db. The attenuation of each section is switched in or out by one flip-flop of a register which contains the desired attenuation in binary form.

A complete vocal tract analog will contain perhaps 20 sections of the form illustrated in Fig. XIV-2, and hence at least 40 coefficients must be supplied by the computer. It is expected that resetting of the coefficient attenuators as often as once every millisecond

might be necessary. The computer, however, would deliver the coefficients in a time sequence, and hence each attenuator will be switching at a different time with respect to the 1-msec cycle.

In this report two problems concerning the performance of this structure are analyzed: The first problem is the question of how well the analog represents the differential equation system describing the electrical model. This question has an aspect of stability and an aspect of error in the natural frequencies of the system. The second problem is the amount of noise introduced by the attenuators on account of switching transients, errors in adjustment, and the sampled and quantized nature of the coefficient signals.

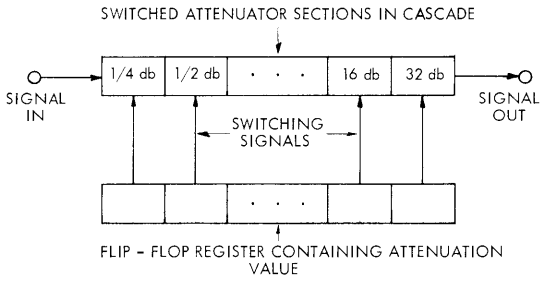


Fig. XIV-3. Principle of a digitally controlled attenuator.

1. Stability and Accuracy

When a set of differential equations is solved on a network of computing amplifiers, the roots of the solution will suffer perturbations, and extraneous roots will be generated by the deviation of amplifier behavior from the ideal mathematical operations that they are supposed to perform The nature of these errors has been discussed by Dow[3] and MacNee.[4] In this report, an analysis is given which is particularly adapted to the design of a dynamic vocal tract analog and should clarify the principles involved.

We suppose that an arbitrary system of linear differential equations is being solved by a configuration of summers, inverters, and identical integrators, in which only the integrators fail to have ideal frequency characteristics. A dynamic vocal tract model constructed from sections of the type shown in Fig. XIV-2b will meet this assumption if the controlled attenuators have perfect frequency response.

Consider the system of differential equations after transformation to the frequency domain through substitution of the complex-frequency variable s for time differentiation and division of each equation by the highest power of s that it contains. In using identical nonideal operational amplifiers for the integration operations, one is effectively replacing appearances of $\frac{1}{s}$ in the transformed equations by the transfer function H(s)

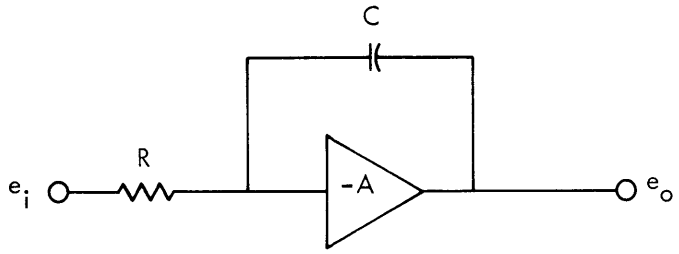


Figure XIV-4.

Conventional arrangement for analog integration.

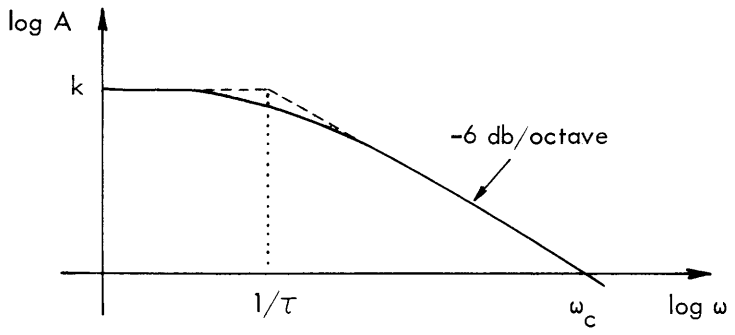
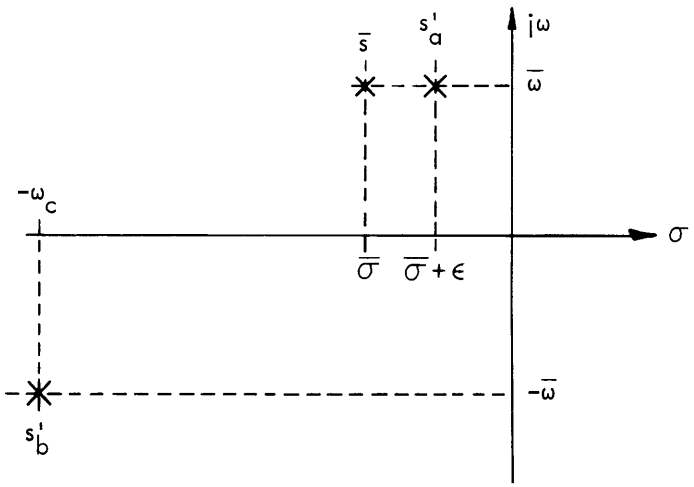


Fig. XIV-5. Frequency dependence of operational amplifier gain.

Figure XIV-6.

Perturbation of a natural frequency by nonideal integration.



of the nonideal integrator. The result of the replacement is a set of frequency domain equations for the analog model in terms of a new complex-frequency variable s'.

$$\frac{1}{s} = H(s'). \tag{1}$$

Suppose that the original differential equation system has roots or natural frequencies

$$s = s_1, s_2, \ldots, s_n. \tag{2}$$

Then the analog system with nonideal integrators will have roots for every value of s' which yields one of the original natural frequencies (2) when substituted in (1). If the transfer function H(s) is nearly $\frac{1}{s}$ in the vicinity of a root s_i of the equation system, then the analog model will have a root s_i' very near s_i. Since (1) may have a number of solutions for any particular value of s, the model will also have extraneous roots. In the following analysis the transformation (1) is constructed for typical amplifier behavior, and the relation between roots of the equation system and roots of the analog model is investigated.

The standard configuration for a practical analog integrator is shown in Fig. XIV-4. The gain A of a practical operational amplifier is well represented by the curve in Fig. XIV-5.

$$A(s) = \frac{k}{1 + \tau s}, \tag{3}$$

where k is the dc gain, τ is the time constant of the principal frequency break point, and $\omega_o = k/\tau$ is the unity gain crossover frequency. The transfer function of the practical integrator is found to be

$$G(s) = \frac{E_o(s)}{E_i(s)} = \frac{\omega_c}{RC\left[s^2 + \left(\frac{1+k}{\tau} + \frac{1}{RC}\right)s + \frac{1}{RC\tau}\right]} = \frac{\omega_c}{RC(s-s_1)(s-s_2)}, \tag{4}$$

where

$$s_1, s_2 = \frac{-b \pm \sqrt{b^2 - 4c}}{2} \tag{5}$$

and, in turn,

$$b = \left(\frac{1+k}{\tau} + \frac{1}{RC}\right); \qquad c = \frac{1}{RC\tau}. \tag{6}$$

For satisfactory performance as an integrator, it is necessary that $\omega_c \gg \frac{1}{RC}$, and therefore $b^2 \gg 4c$. Hence

$$s_1 = -b = -\left(\frac{1+k}{\tau} + \frac{1}{RC}\right) \approx -\omega_c \tag{7}$$

$$s_2 = \frac{-b - b\sqrt{1 - \frac{4c}{b^2}}}{2} \approx -\frac{c}{b}$$

$$s_2 \approx \frac{-1}{kRC + \tau}. \tag{8}$$

The frequency range utilized for computation is the range for which

$$|s_2| \ll |s| \ll |s_1|,$$

and in this range

$$G(s) \approx \frac{1}{RCs}.$$

Therefore the appropriate function for use in (1) is

$$H(s) = RCG(s) = \frac{\omega_c}{(s-s_1)(s-s_2)}, \tag{9}$$

and the conversion from the equation system to the analog model is described by the change of variable

$$s = \frac{(s'-s_1)(s'-s_2)}{\omega_c}. \tag{10}$$

According to this substitution, each natural frequency $s = \bar{s}$ of the equation system becomes a pair of natural frequencies in the analog. To find the relation between roots of the model and those of the equation system, we solve (10) for s' in terms of $\bar{s}$. Setting $\bar{s} = \bar{\sigma} + j\bar{\omega}$ and using the approximations $s_1 \gg s_2$ and $s_1 \approx \omega_c$, we find

$$s'_a = \sigma'_a + j\omega'_a \approx \bar{s} + \frac{\bar{\omega}^2}{\omega_c} - \frac{1}{kRC + \tau} \tag{11}$$

and

$$s'_b = \sigma'_b + j\omega'_b \approx -\omega_c - j\bar{\omega}. \tag{12}$$

As shown in Fig. XIV-6, s'_a is the original natural frequency with its real part perturbed by an amount

$$\epsilon = \frac{\bar{\omega}^2}{\omega_c} - \frac{1}{kRC + \tau}. \tag{13}$$

The root s_b' is extraneous, but its large negative real part prevents its affecting the solution seriously. These results agree with those stated by MacNee[4] and with experimental findings of Hills[5] and the present authors.

The shift of the important root has two components. The positive one, which could lead to instability if excessive, is proportional to $1/\omega_c$. The negative component is approximately proportional to $\frac{1}{\tau} = \frac{\omega_c}{k}$ and leads to exaggerated damping. The minimization of both effects places conflicting requirements on ω_c, and a judicious compromise will be necessary to ensure optimum results.

2. Noise

Noise introduced by the digitally controlled attenuator component can be divided into two categories; additive noise – that which is present in the absence of a signal – and multiplicative noise – that which is proportional in intensity to the signal. The multiplicative noise arises from the sampling and quantization processes inherent in the digital origin of the coefficient signals. The additive noise is introduced primarily by switching transients and errors in adjustment.

Introduced sampling and quantization noise can be analyzed with the help of the model in Fig. XIV-7. In the model the sampling operation precedes quantization for convenience of analysis. Although the quantization steps are actually proportional to coefficient

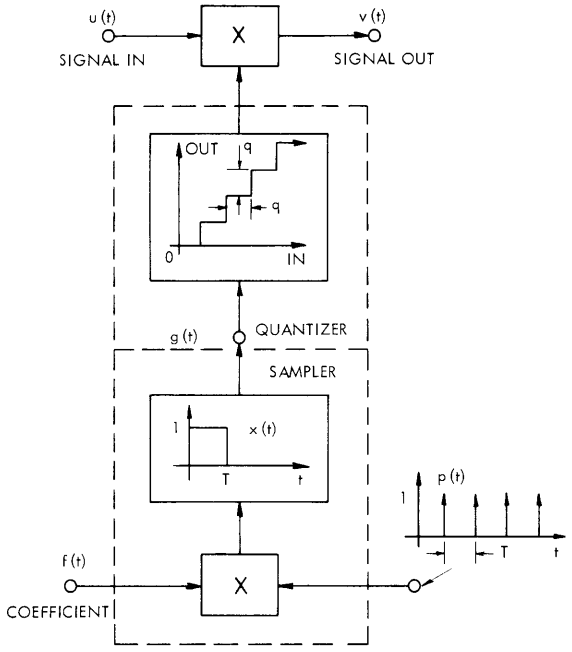


Fig. XIV-7. Model of digital attenuator for analysis of sampling and quantization noise.

values, the use of uniform steps is adaquate, for the coefficients have slowly varying values.

Note that if we fix attention on a particular control waveform f(t), the model is a linear time-varying system with regard to its effect on the signal u(t). Therefore, the effect of the control waveform on a complex signal can be determined from the effect on a sinusoid through the principle of superposition.

Consider, then, $u(t) = \cos \omega t$. The desired output is $w(t) = u(t)[1+f(t)]$, and the actual output is the desired output plus noise.

$$v(t) = w(t) + n(t).$$

Since f(t) is a slowly varying function, its spectrum $F(\omega)$ is concentrated at low frequencies. The spectrum $G(\omega)$ of the sampler output is $F(\omega)$ replicated at intervals of $2\pi/T$ along the radian frequency axis and multiplied by the squared magnitude system function

$$T^2 \frac{\sin^2 \frac{\omega T}{2}}{\left(\frac{\omega T}{2}\right)^2}$$

of the boxcar circuit. This computation is illustrated in Fig. XIV-8. The low-frequency component of $|G(\omega)|^2$ is the desired signal with high frequencies cut down slightly by the

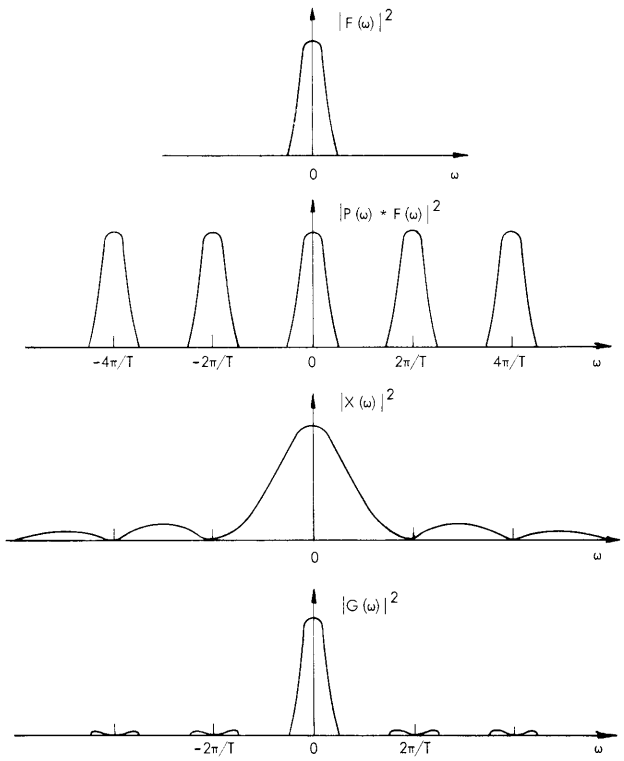


Fig. XIV-8. Computation of power spectrum of sampled coefficient signal.

sin x/x function. The degradation will be small if the bandwidth of f(t) is small compared with $2\pi/T$. The other components of $G(\omega)$ are noise.

The quantization operation adds a noise component that has the form of a rectangular wave with random amplitude values in the range from $-\frac{q}{2}$ to $+\frac{q}{2}$, as shown in Fig. XIV-9a. The statistical theory of quantization[6] shows that these amplitude values may be regarded as independent samples from a uniform distribution over this range, if the quantization box is small enough so that adjacent samples are likely to differ by more than one step. Experience with quantized systems has shown that this condition on box size does not have to be well satisfied to achieve apparent independence of the quantization noise. Under these conditions, the power spectrum of the quantization noise will be the $\sin^2 x/x^2$ curve shown in Fig. XIV-9b. The sampling and quantization noise components reach the attenuator output only through modulation of the signal. For a sinusoidal signal, the noise spectra appear at the output shifted by the signal frequency. For a complex signal, the resultant noise will be the superposition of these noise spectra with various frequency shifts according to the frequency content of the signal.

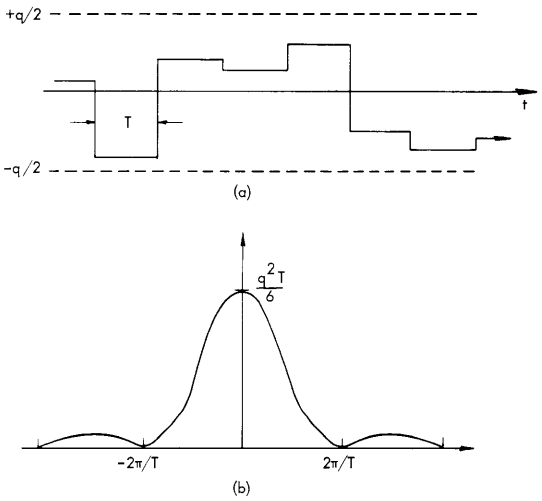


Fig. XIV-9. Form of quantization noise as a random-amplitude square wave and its power spectrum.

The additive component of noise originating from switching transients takes the form of a train of short pulses having pseudorandom amplitudes and a uniform time spacing T. Such a waveform has a uniform power density in the audio-frequency range if the pulses have independent amplitudes. An error in adjustment of an attenuator section will cause some of the switching signal to appear at the attenuator output. If each has an unrelated error of this type, the corresponding noise component will be a random-amplitude square wave having the form of spectrum shown in Fig. XIV-9b.

These sources of noise in the attenuator output make a careful choice of parameters necessary in the design of the attenuator. At present, an attenuator design is being evaluated with respect to noise contributions and their effect on the performance of an analog vocal tract model.

J. B. Dennis, E. C. Whitman

References

1. E. C. Whitman, An Improved Dynamic Analog of the Vocal Tract, S.M. Thesis, Department of Electrical Engineering, M.I.T., January 1963.

2. E. C. Whitman, A transistorized articulatory speech synthesizer, Quarterly Progress Report No. 68, Research Laboratory of Electronics, M.I.T., January 15, 1963, pp. 164-167.

3. P. C. Dow, An analysis of certain errors in electronic differential analyzers, IRE Trans., Vol.EC-6, pp. 255-260, December 1957.

4. A. B. MacNee, Some limitations on the accuracy of electronic differential analyzers, Proc. IRE 40, 304-308 (1952).

5. F. B. Hills, Application Studies of Combined Analog-Digital Computation Techniques, Report ESL-FR-165, Electronic Systems Laboratory, M.I.T., February 1963.

6. B. Widrow, A study of rough amplitude quantization by means of Nyquist sampling theory, IRE Trans., Vol. PGCT-3, No. 4, pp. 266-276, December 1956.